

Evaluating the potential for statistical decadal predictions of sea surface temperatures with a perfect model approach:

Supplementary Information

ED HAWKINS*, JON ROBSON, ROWAN SUTTON, DOUG SMITH, NOEL KEENLYSIDE

This supplementary information augments the findings of the main paper by showing the global versions of many of the figures, and by including the skill of the prediction methods that were not shown in the main paper. Additionally, we include a brief discussion comparing the leading EOFs of the observations and GCMs. Finally, we discuss the skill of SST predictions in operational decadal prediction systems.

Potential predictability

Fig. S1 is the global version of Fig. 1 and shows the potential predictability (PP) estimates for the two GCMs and observations outside the Atlantic. Note particularly that Fig. S1b shows PP using the raw observations (top) and the observations detrended with a spline as in the main text (bottom).

It can be seen that much of the Atlantic has the highest levels of potential predictability, although the Pacific is also highlighted. Although the presence of potential predictability is not *necessarily* an indicator of actual predictability, this motivates considering predictions focussed on the Pacific in more detail in future work to establish whether this potential predictability can be realised.

Leading EOFs of models and observations

In the LIM analysis in the main paper, EOFs are used to build the statistical model. Fig. S2 shows the leading EOFs of the observations (HadISST; Rayner et al. 2003) and the control runs of the two GCMs considered in this study. Note that in the LIM analysis, the EOFs are estimated for each 140 year training period separately, but here we use the whole control run.

*E-mail: e.hawkins@reading.ac.uk

The leading EOF of HadISST shows an overall warming, likely associated with the radiatively forced signal. The leading EOF of both GCMs shows a pattern with opposite signs across the equator, which also resembles EOF2 of the observations. The second EOF of both GCMs shows a similarly signed signal either side of the equator, and this pattern is not clearly seen in EOF3 of the observations. This is not particularly surprising as EOFs will not separate the natural and anthropogenic components cleanly.

Prediction skill

Figs. S3-S6 are the global versions of Figs. 2-5, showing the skill measures over the entire global domain for all prediction types, and they highlight the other regions where some skill may be found. In particular, both the global LIM and CA methods show high levels of skill in the Indian (HadCM3) and Pacific (HadGEM1) basins, although only for lead times up to 2 years. From a comparison of the global and Atlantic-only predictions it can be seen that using the global SSTs is often superior in year 1 in the tropical Atlantic, likely related to predicting aspects of ENSO in the Pacific.

Comparison of operational decadal prediction systems

We now qualitatively compare the potential skill of the statistical approaches with the actual skill found in two operational GCM-based decadal prediction systems, DePreSys (Smith et al. 2007) (hereafter S07), which is based on HadCM3, and Keenlyside et al. (2008) (hereafter K08), which uses the ECHAM5/MPI-OM climate model. In these decadal predictions, skill comes from both predicting the trend and the variability around the trend. We also use the companion simulations which only predict the trend (denoted as NoAssim for S07 and ‘20th C’ for K08). Note that we do not consider any forecast bias correction¹.

For DePreSys (S07), we use the original retrospective forecasts (or hindcasts) which are started every season from 1981-2001, and we consider the skill of the ensemble mean of all 4 members for each prediction. Both the predictions and the NoAssim simulations do not include knowledge of volcanic eruptions that occur later in each hindcast. The predictions from K08 are started every 5 years from 1955-2005, and we consider the skill of the ensemble mean of all 3 members for each prediction. Again, the initialised

¹It is possible that the skill measures used are sensitive to the bias correction and this will be explored in future work.

predictions do not include future volcanic eruptions, but the ‘20th C’ simulations have the volcanic forcing prescribed. The other major difference between the two systems is that K08 only assimilate historical SSTs into the model, whereas S07 nudge to an analysis of temperature and salinity at all depths (Smith and Murphy 2007). Table 1 summarises the experimental configurations. When estimating the skill in these predictions below, the SST anomalies in the hindcasts are compared to observed anomalies relative to a chosen climatological period (1955-2004 for K08 and 1951-2006 for DePreSys [following Robson 2010]).

Anomaly correlation

There is no simple decomposition of the correlations of such an initialised and uninitialised decadal prediction system. However, Fig. S7a shows the SST correlation skill of the predictions from S07 for NoAssim (top row), DePreSys (middle row) and the difference (bottom row). The initialised system, DePreSys, generally outperforms NoAssim for year 1, especially in the tropics and in the far North Atlantic. At longer lead times there is additional skill only in the far North Atlantic.

Although a direct quantitative comparison between DePreSys and Fig. S3 is not appropriate, it is interesting to note that the initialised system does outperform the uninitialised system in many regions for year 1, and some regions for longer lead times. As demonstrated, the statistical methods are also providing additional skill over a climatological forecast in some regions, suggesting that they could match the skill of the GCM-based predictions, and this motivates their further development.

Fig. S8a shows the equivalent SST correlation skill from K08, again demonstrating increased predictive ability for the first year in most regions, and the north-east Atlantic at longer lead times. The correlation skill in years 6-10 is still above 0.8 over a large region, however many regions see a reduction in skill with initialisation. Note that the 95% confidence level for these correlations, assuming 9 independent predictions, is $r \approx 0.7$.

Both systems seem to show the potential to predict ENSO up to 2 years ahead. Beyond 2 years, the main skill outside the Atlantic is found in the east Pacific for K08.

Relative RMS error

As noted in the main paper, it is important to consider more than one measure of skill. For a comparison of the RMS error we consider that,

$$\text{RMS}_{\text{relative}} = \frac{\text{RMS}_{\text{init}}}{\text{RMS}_{\text{uninit}}}, \quad (1)$$

where ‘init’ and ‘uninit’ refer to the initialised and uninitialised predictions respectively. Note that this measure is relatively insensitive to the presence of the same trend in both predictions.

Fig. S7b shows this relative RMS error for DePreSys (S07) and identifies similar regions to the correlations where there is enhanced skill for DePreSys over NoAssim (shown in blue colours), particularly for the far North Atlantic.

Fig. S8b shows the relative RMS error for K08, again demonstrating improved skill for year 1. However, for longer lead times the RMS error is significantly worse than the ‘20th C’ projections for many regions. It may also be noted that regions of high correlation often also have large RMS errors, and this is probably due to the model’s internal variability being too large in this region (see K08).

Overall, it is seen that the initialised GCM based predictions provide some additional skill in predicting the internal variability than the uninitialised simulations, especially on shorter lead times. The trends provide most of the apparent skill on longer lead times.

References

- Keenlyside NS, Latif M, Jungclauss J, Kornbluh L, Roeckner E (2008) Advancing decadal-scale climate prediction in the North Atlantic sector. *Nature* 453: 84–88 doi: 10.1038/nature06921
- Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J Geophys Res* 108: 4407 doi: 10.1029/2002JD002670
- Robson JJ (2010) Understanding the performance of a decadal prediction system Ph.D. thesis University of Reading, UK

Smith DM, Cusack S, Colman AW, Folland CK, Harris GR, Murphy JM (2007) Improved surface temperature prediction for the coming decade from a global climate model. *Science* 317: 796–799 doi:10.1126/science.1139540

Smith DM, Murphy JM (2007) An objective ocean temperature and salinity analysis using covariances from a global climate model. *J Geophys Res* 112: C02022 doi:10.1029/2005JC003172

Table 1: Description of the GCM based decadal prediction systems considered, Smith et al. (2007) and Keenlyside et al. (2008). The two systems are based on different GCMs, consider different historical time periods, have different frequency of start dates, and initialisation strategies for both temperature (T) and salinity (S).

Study	GCM used	Years considered	Frequency	Initialisation strategy
S07	HadCM3	1981-2001	every season	anomalies, full depth T, S
K08	ECHAM5/MPI-OM	1955-2005	every 5 years	anomalies, SST only

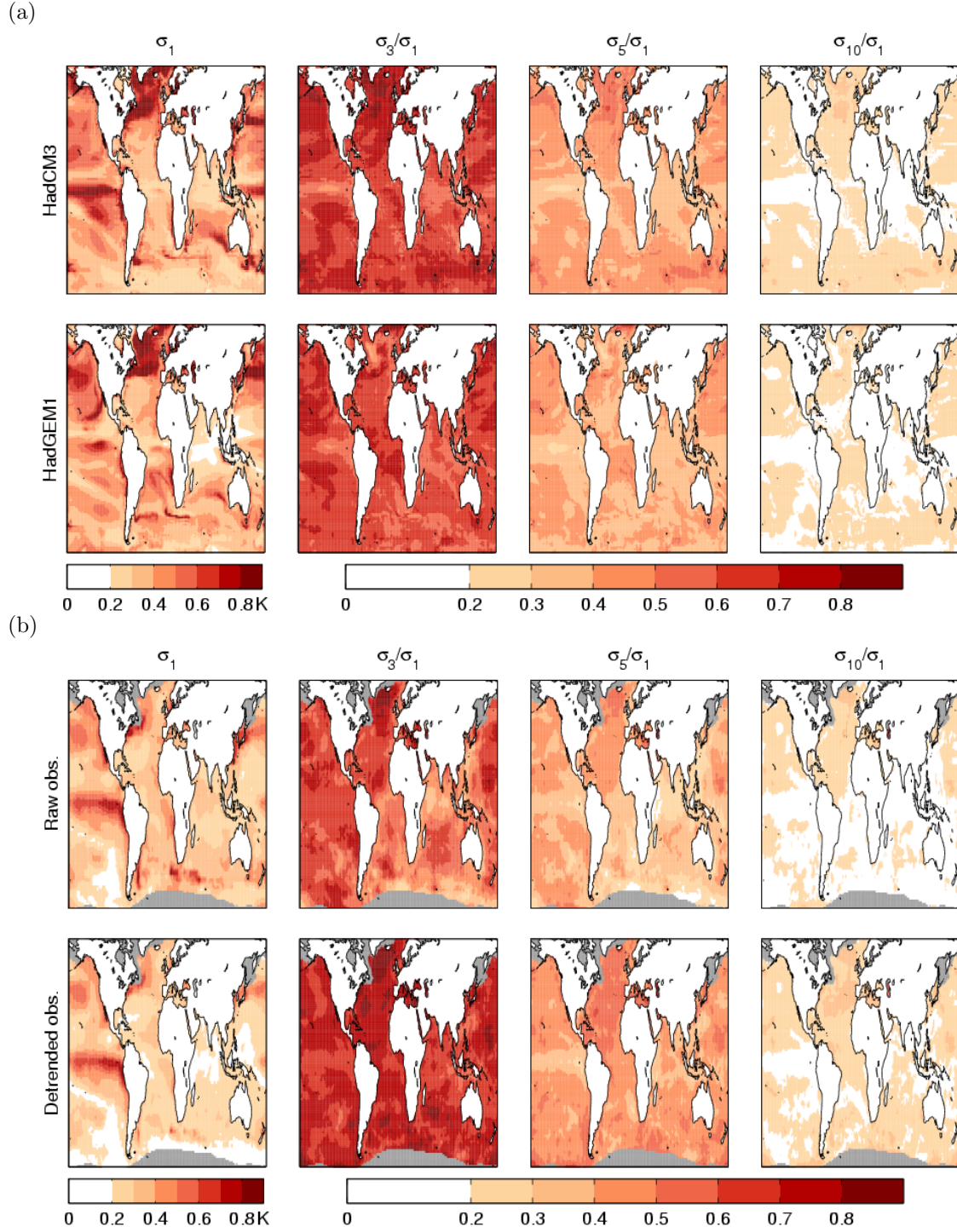


Figure S1: Potential predictability of Atlantic SSTs. Left column: interannual variability of SST. Other columns: ratio of standard deviation of 3, 5 and 10 year means to the interannual standard deviation. (a) Top row: HadCM3. Bottom row: HadGEM1. (b) HadISST observations from 1870-2009. Top row: raw observations. Bottom row: detrended observations, using cubic spline. Regions where sea ice precludes analysis are shown in grey.

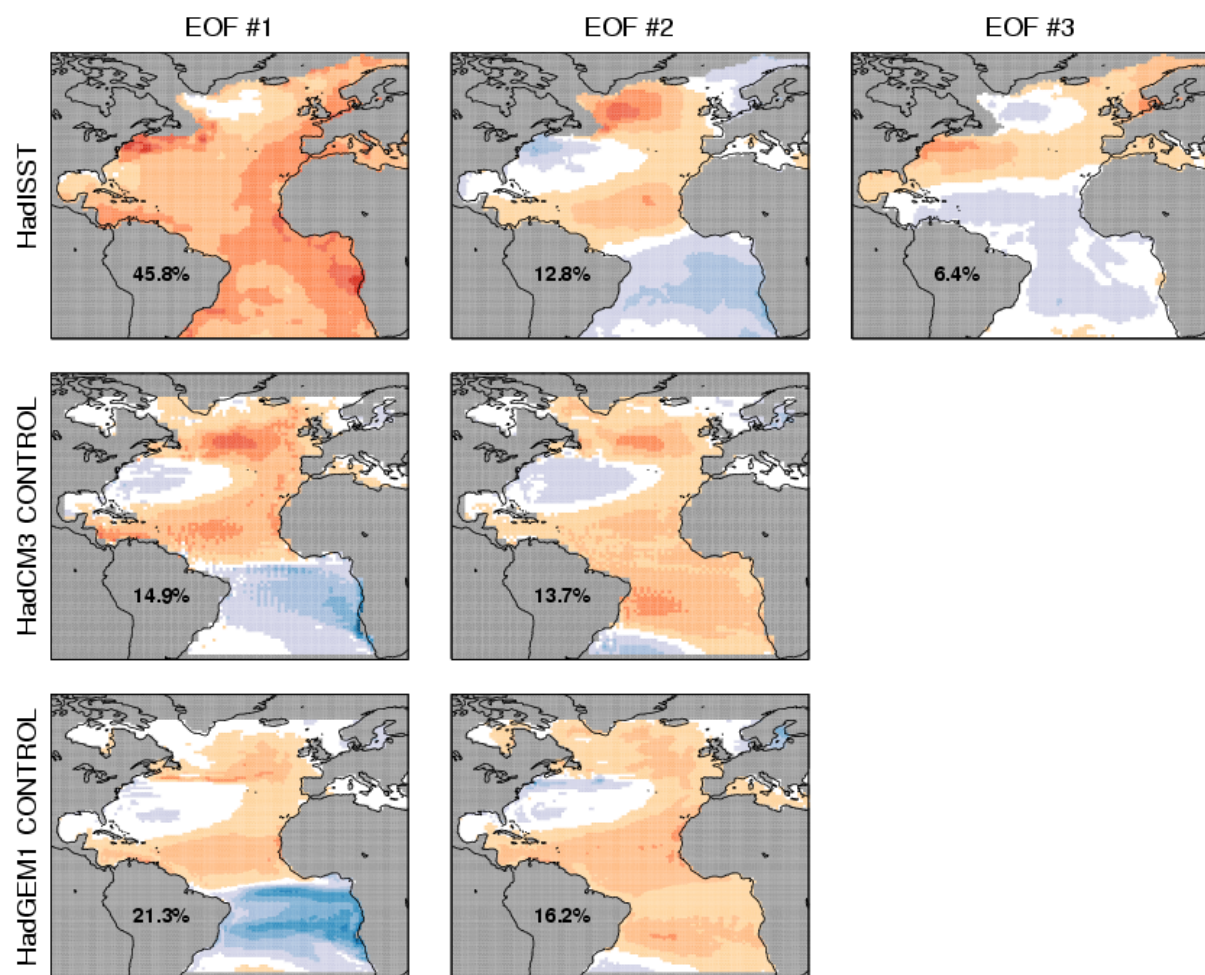


Figure S2: The leading correlation EOFs of the observations (HadISST, top) and the control runs of HadCM3 and HadGEM1.

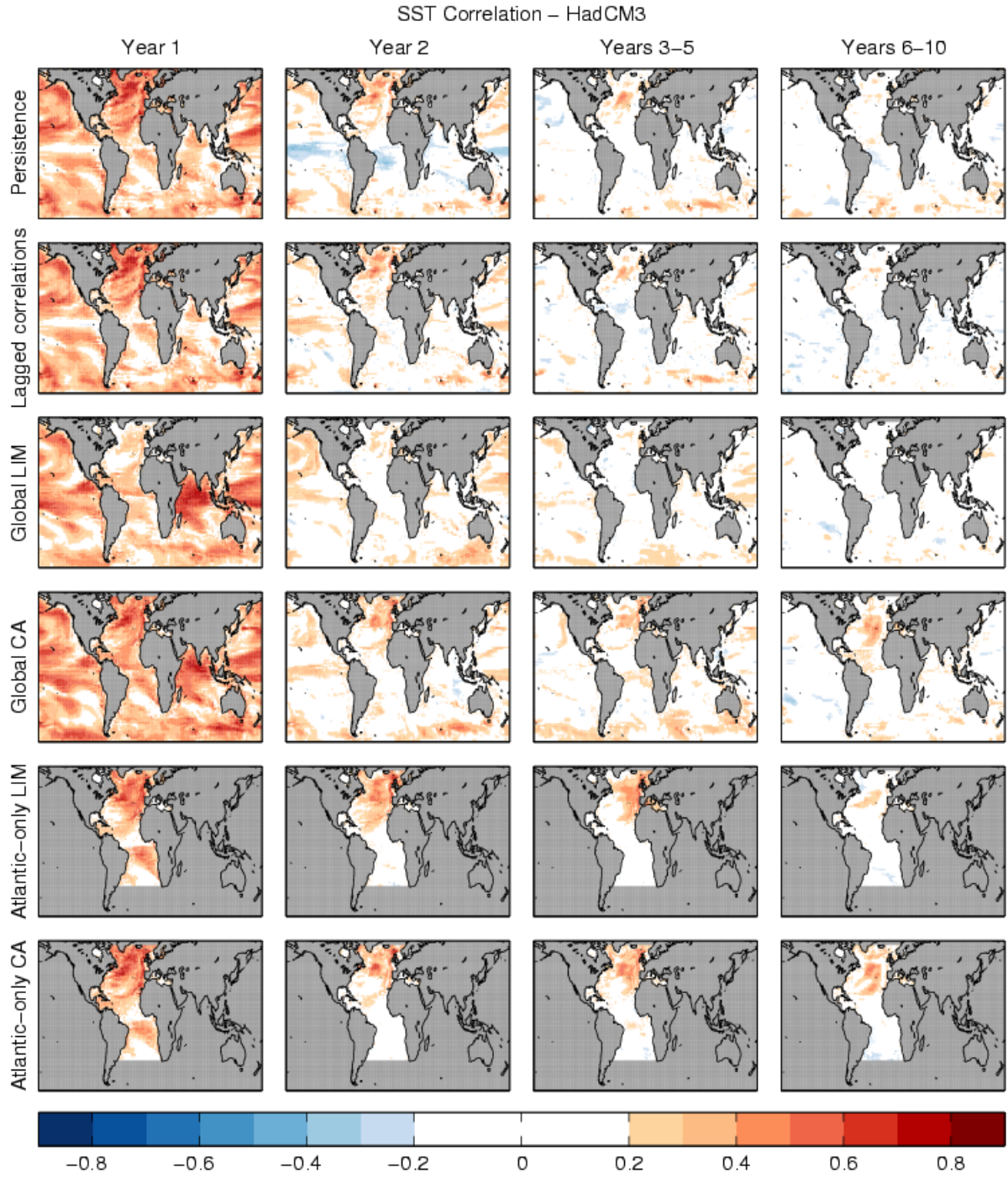


Figure S3: SST anomaly correlation skill for HadCM3, for persistence, lagged correlations, Linear Inverse Modelling (LIM) and Constructed Analogue (CA) predictions. The LIM and CA methods are repeated for both Global and Atlantic domains. The grey regions are masked out of the analysis.

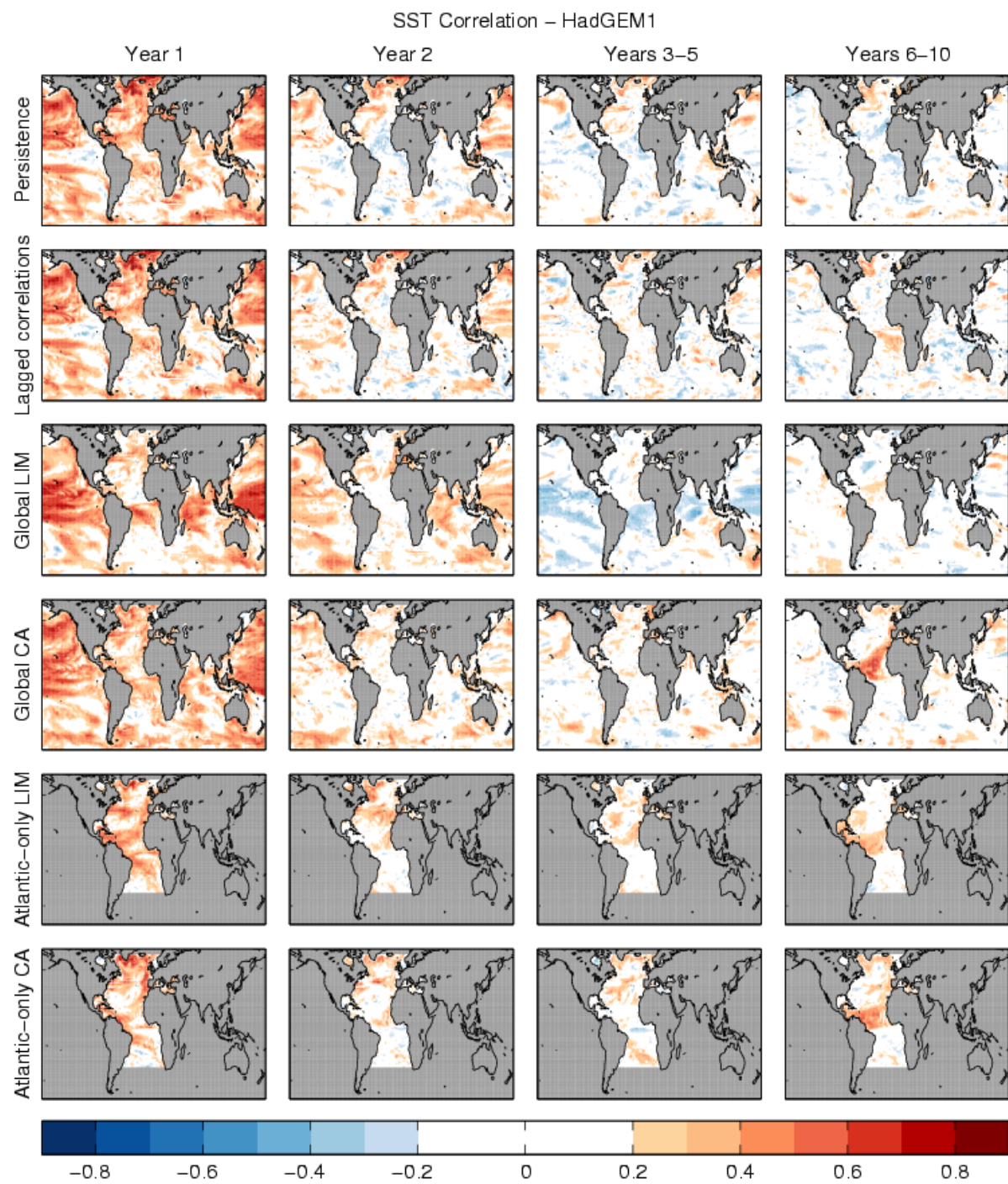


Figure S4: Same as Fig. S3 for HadGEM1.

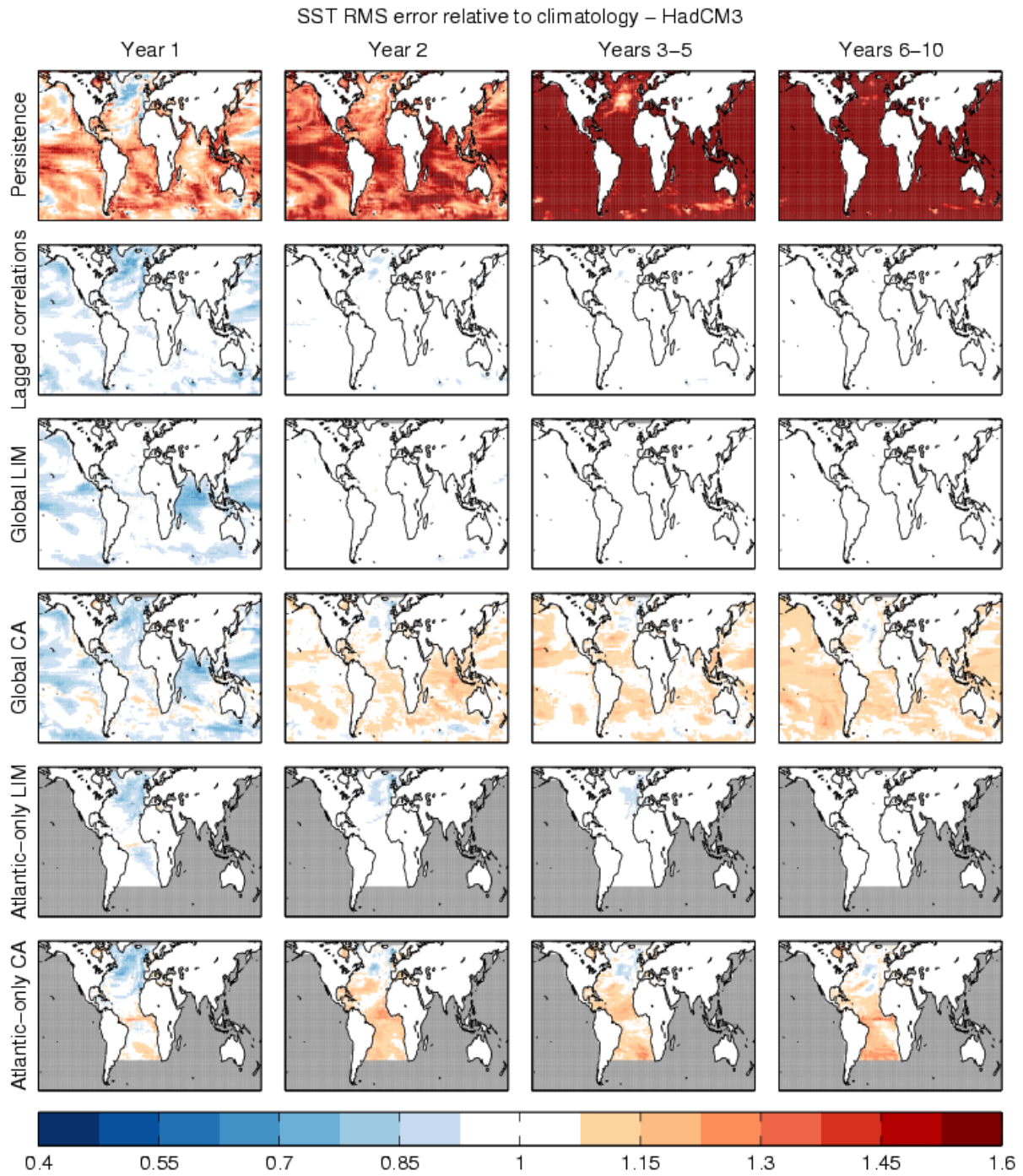


Figure S5: Root mean square error relative to the climatological RMS error for HadCM3, for persistence, lagged correlations, LIM and CA predictions. The LIM and CA methods are repeated for both Global and Atlantic domains. The grey regions are masked out of the analysis.

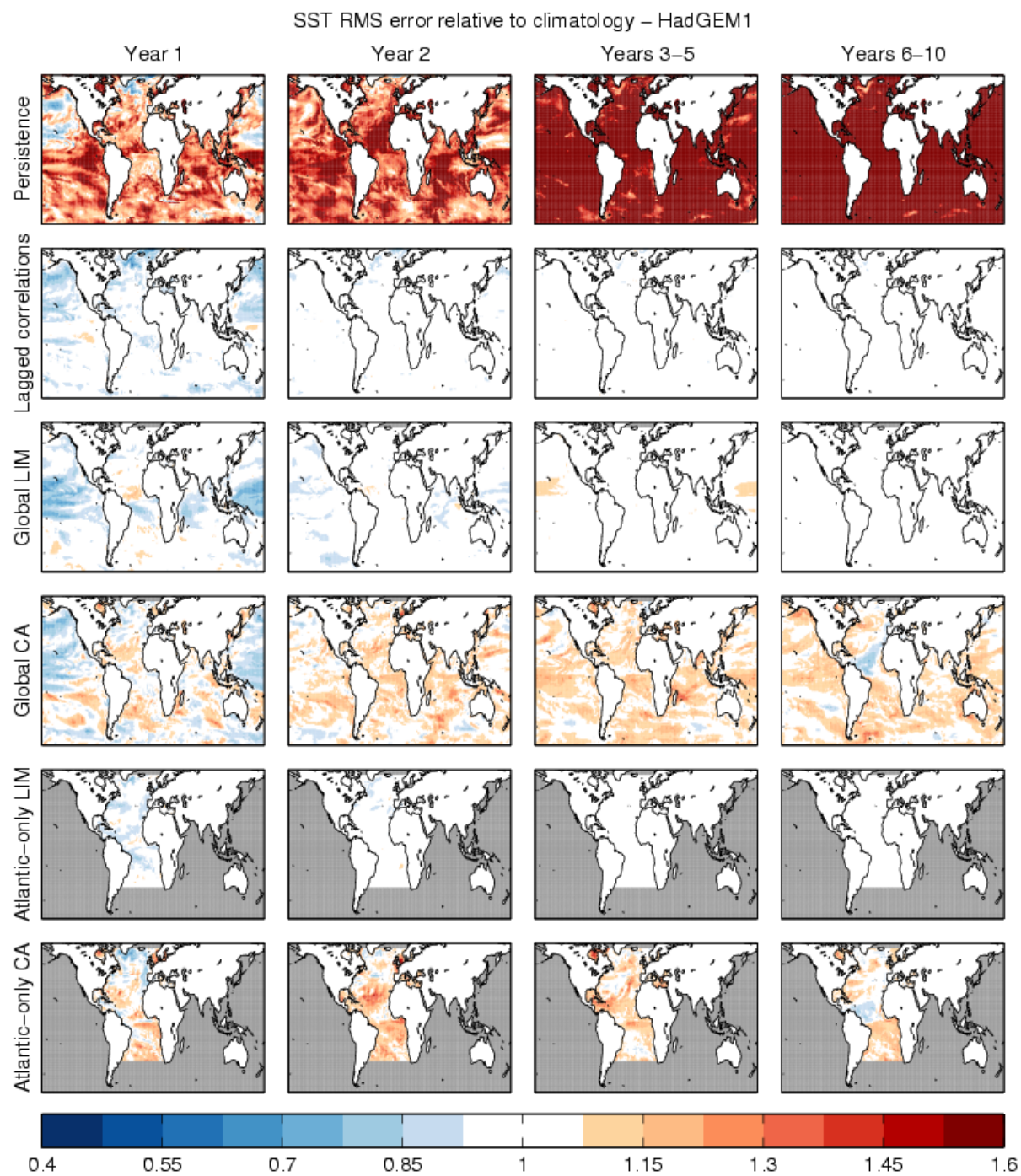
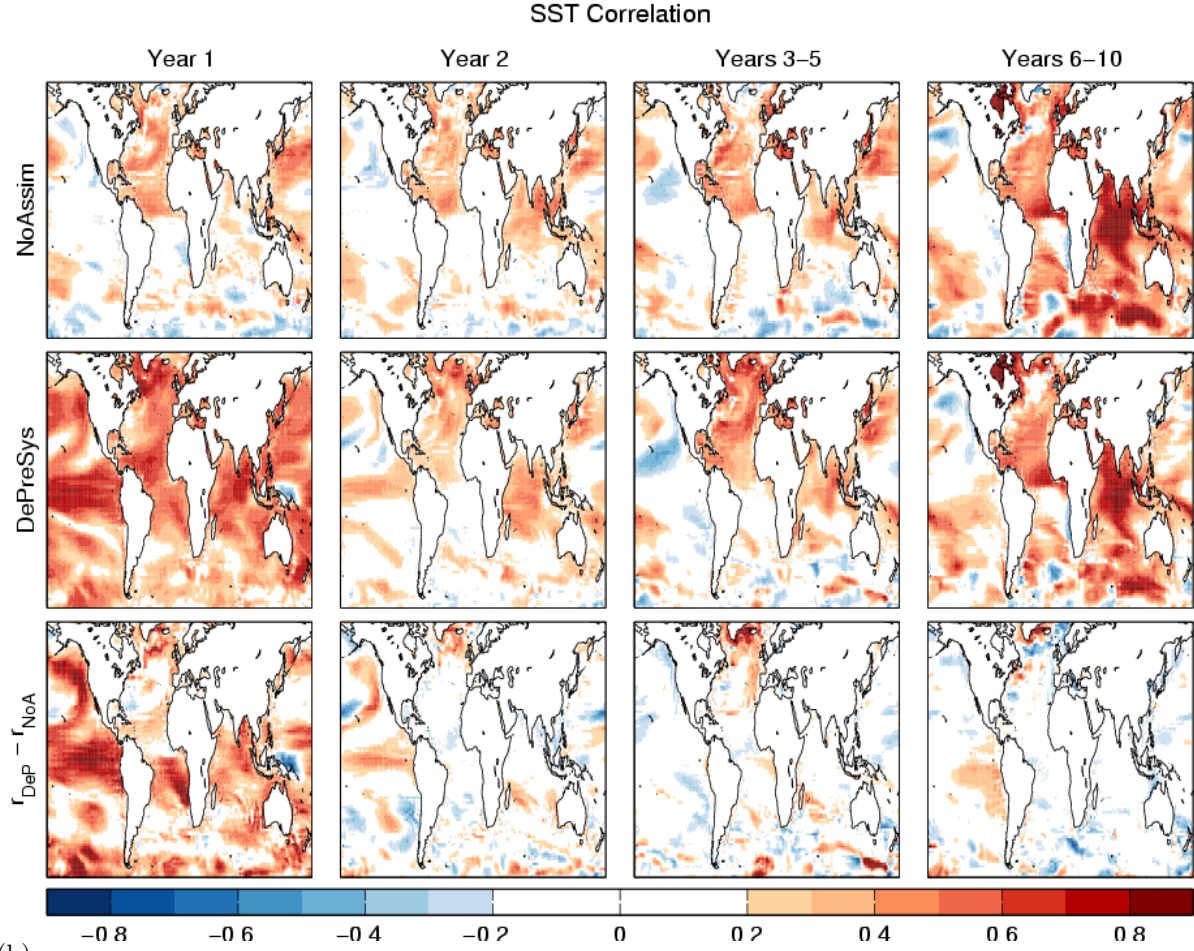


Figure S6: Same as Fig. S5 for HadGEM1.

(a)



(b)

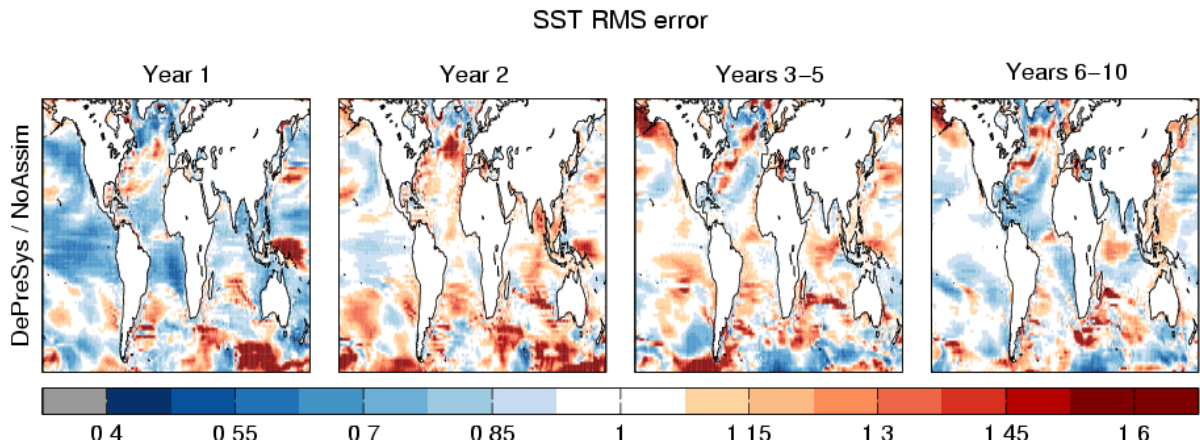
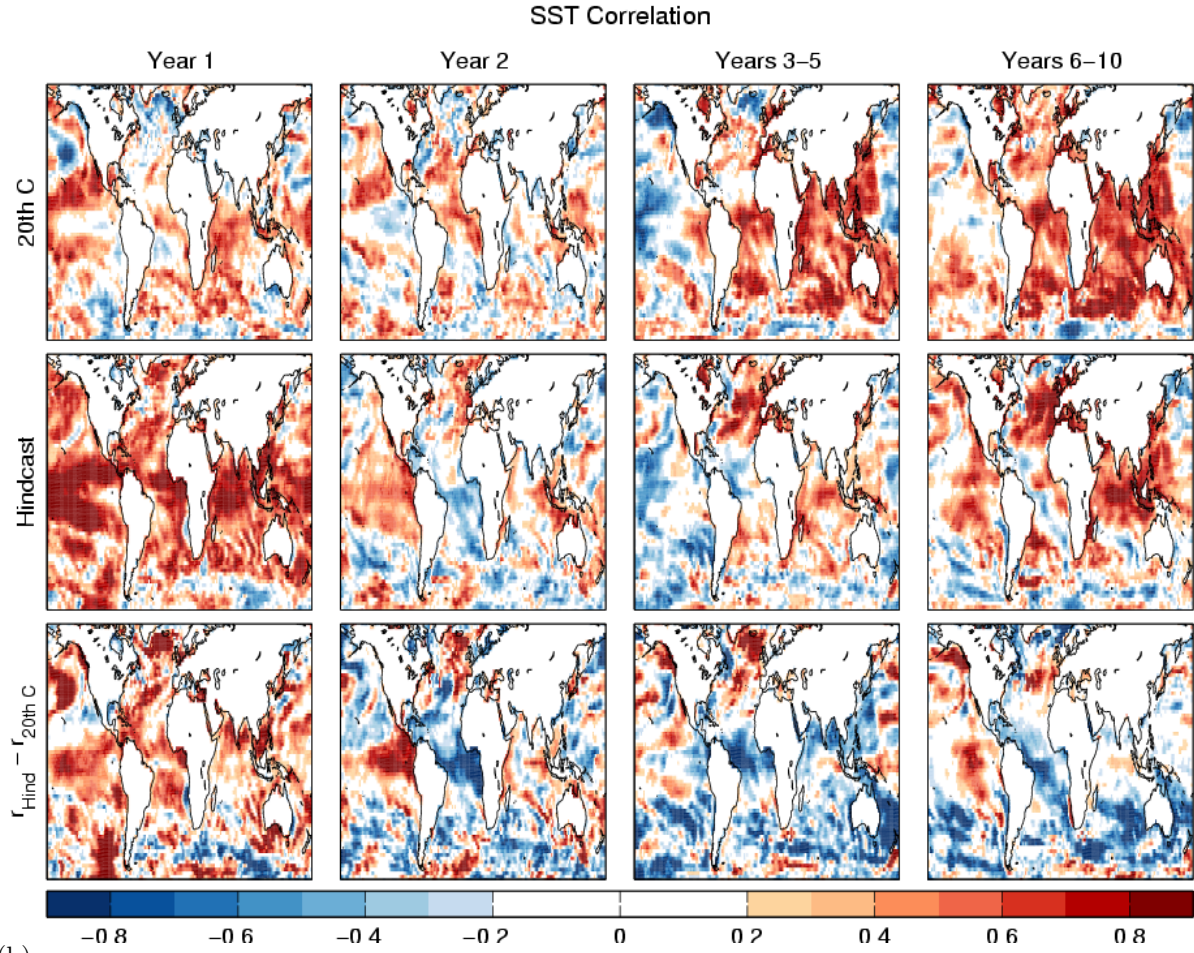


Figure S7: Prediction skill of an operational GCM-based decadal prediction system (DePreSys; Smith et al. 2007) for years 1981-2001. (a) Correlation skill for the uninitialised ensemble (NoAssim), initialised ensemble (DePreSys) and the difference. (b) RMS error of DePreSys relative to NoAssim.

(a)



(b)

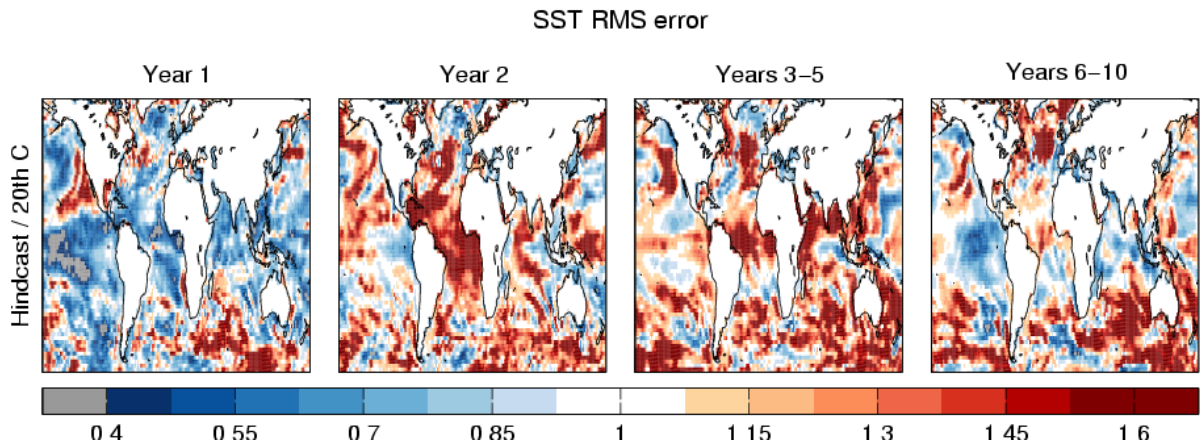


Figure S8: Prediction skill of an operational GCM-based decadal prediction system (Keenlyside et al. 2008) for years 1955–2005. (a) Correlation skill for the uninitialised ensemble (20th C), initialised ensemble (Hindcast) and the difference. (b) RMS error of Hindcast relative to 20th C.